# ASSET: A Robust Algorithm for the Automated Segmentation and Standardization of Early Caenorhabditis elegans Embryos

Blanchoud, Simon ; Budirahardja, Yemima ; Naef, Felix ; Gönczy, Pierre

Abstract: The early Caenorhabditis elegans embryo is an attractive model to investigate evolutionarily conserved cellular mechanisms. However, there is a paucity of automated methods to gather quantitative information with subcellular precision in this system. We developed ASSET (Algorithm for the Segmentation and the Standardization of C. elegans Time-lapse recordings) to fill this need. ASSET automatically detects theeggshell and the cell cortex from DIC time-lapse recordings of live one-cell-stage embryos and can also track subcellular structures using fluorescent time-lapse microscopy. Importantly, ASSET standardizes the data into an absolute coordinate system to allow robust quantitative comparisons between embryos. We illustrate how ASSET can efficiently gather quantitative data on the motion of centrosomes and precisely track cortical invaginations, revealing hitherto unnoticed differences between wild-type and saps-1(RNAi) embryos. In summary, we establish ASSET as a novel tool for the efficient quantification and standardization of images from early C. elegans embryos.

DOI: https://doi.org/10.1002/dvdy.22486

Developmental Dynamics

# ASSET: a robust algorithm for the automated segmentation and standardization of early C. elegans embryos



| | |
|---|---|
| Journal: | *Developmental Dynamics* |
| Manuscript ID: | DVDY-10-0345.R1 |
| Wiley - Manuscript type: | Techniques |
| Date Submitted by the Author: | n/a |
| Complete List of Authors: | Blanchoud, Simon; EPFL, ISREC<br>Budirahardja, Yemima; EPFL, ISREC<br>Naef, Félix; EPFL, IBI<br>Gönczy, Pierre; EPFL, ISREC |
| Keywords: | C. elegans, embryo, DIC microscopy, image analysis, automation, segmentation, quantification, standardization |



Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

movieS1.mov
movieS2.mov
movieS3.mov
movieS4.mov

# ASSET: a robust algorithm for the automated segmentation and standardization of early *C. elegans* embryos

Simon Blanchoud[*,**], Yemima Budirahardja[*,§], Félix Naef[**,#], Pierre Gönczy[*,#]

[*] Swiss Institute for Experimental Cancer Research (ISREC), School of Life Sciences, Swiss Federal Institute of Technology (EPFL), Lausanne CH-1015, Switzerland.
Phone number: +41216930710

**The Institute of Bioengineering (IBI), School of Life Sciences, Swiss Federal Institute of Technology (EPFL), Lausanne CH-1015, Switzerland.
Phone number: +41216931621

[§] Current address: Program in Neuroscience and Behavioral Disorders, Duke-NUS Graduate Medical School, Singapore 169857, Singapore

[#] These authors contributed equally.

Corresponding authors:
pierre.gonczy@epfl.ch
felix.naef@epfl.ch

Running title: An algorithm for the segmentation and the standardization of *C. elegans* embryos

Keywords: *C. elegans*, embryo, DIC microscopy, image analysis, automation, segmentation, quantification, standardization

Grant Sponsors: "Josy Marty – Echec au Cancer de la Broye" to SB, SystemsX.ch IPP 08-027 to YB, SystemsX.ch IPhD 51PH_0125979 to SB, Swiss National Science Foundation grant 3100A0-122500/1 to PG.

## Abstract

The early *Caenorhabditis elegans* embryo is an attractive model to investigate evolutionarily conserved cellular mechanisms. However, there is a paucity of automated methods to gather quantitative information with subcellular precision in this system. We developed ASSET (Algorithm for the Segmentation and the Standardization of *C. elegans* Time-lapse recordings) to fill this need. ASSET automatically detects the eggshell and the cell cortex from DIC time-lapse recordings of live one-cell stage embryos and can also track subcellular structures using fluorescent time-lapse microscopy. Importantly, ASSET standardizes the data into an absolute coordinate system to allow robust quantitative comparisons between embryos. We illustrate how ASSET can efficiently gather quantitative data on the motion of centrosomes and precisely track cortical invaginations, revealing hitherto unnoticed differences between wild-type and *saps-1(RNAi)* embryos. In summary, we establish ASSET as novel tool for the efficient quantification and standardization of images from early *C. elegans* embryos.

## Introduction

Quantitative cell and developmental biology have benefited tremendously from improvements in imaging techniques. However, accurate quantification of image data is typically slow and repetitive. Therefore, developing automated solutions that result in robust data extraction from images of noisy biological processes is a long sought-after goal. Such reliable quantification is also a pre-requisite for developing and testing quantitative mathematical models.

The early embryo of the nematode *C. elegans* has proven particularly well suited to study fundamental cellular processes in a developing organism (reviewed in Oegema and Hyman, 2006). These processes can be monitored with time-lapse recordings of live embryos using standard differential interference contrast (DIC) microscopy. In addition, imaging of given fusion proteins or cellular compartments can be achieved using time-lapse fluorescence microscopy. The embryo develops inside a ~50µm by ~30µm elliptical rigid eggshell. This large size provides excellent spatial resolution. In addition, the rapid duration of the first cell cycle (~20min) allows high temporal resolution. Moreover, forward genetic and RNAi-based functional genomic screens have led to the systematic identification of evolutionarily conserved components required for fundamental cellular processes (reviewed in Oegema and Hyman, 2006). Quantitative analysis of gene expression patterns at the single cell level can be achieved in fixed specimens of *C. elegans* (Bao et al., 2006; Long et al., 2009) and existing algorithms can quantify specific processes in live embryos (see for instance Hamahashi et al., 2005; Goulding et al., 2007; Jaensch et al., 2010). However, there is currently no algorithm that combines image segmentation with standardization to allow quantitative analysis of distinct cellular processes in a common framework.

The first cell cycle of *C. elegans* embryos is characterized by a stereotyped sequence of events that can be observed using time-lapse DIC microscopy (Movie S1). Shortly after meiosis, the female and male pronuclei form in the presumptive embryo anterior and

posterior, respectively, and are visible because they exclude the yolk granules that fill the cytoplasm (Fig. 1A, 1E). Initially, characteristic ruffles are apparent throughout the surface of the embryo, reflecting actomyosin-based contractions of the cell cortex located just below the plasma membrane. Thereafter, the sperm-derived centrosome initiates polarity establishment along the anteroposterior (AP) embryonic axis (reviewed in Gönczy, 2008). As a result, non-contractile cortex expands from the posterior side and the contractile cortex becomes gradually restricted to the embryo anterior (Fig. 1B, 1F). At this stage, a prominent pseudocleavage furrow forms towards the center of the embryo. Concomitantly, the two pronuclei migrate towards each other, meet in the embryo posterior, after which they move together to the cell center, along with the associated centrosomes (Fig. 1C, 1G). This is followed by entry into mitosis, which is visible by breakdown of the nuclear envelopes and assembly of a bipolar spindle. Unequal pulling forces act on the spindle poles during anaphase, such that the cleavage furrow is positioned slightly to the posterior at the onset of cytokinesis, which results in the cleavage of the one-cell stage embryo into two unequal daughter cells.

Quantifying the changes that occur during this developmental sequence of processes, in particular those taking place at the dynamic cell cortex, would require tedious manual tracking of a large number of time-lapse recordings. The segmentation, i.e. the partitioning of the image into various domains based on specific criteria, of even a single frame is a time-consuming task that is not scalable. Consequently, we set out to develop an automated image analysis software to track cellular processes in early *C. elegans* embryos.

Many scientists working with the early embryo of *C. elegans* use time-lapse DIC recordings to monitor cellular processes. Therefore, we decided to develop a novel image analysis pipeline based on the segmentation of DIC images termed ASSET (Algorithm for the Segmentation and the Standardization of *C. elegans* Time-lapse recordings). As we aim at replacing manual segmentation, we compared the performance of the algorithm with manual tracking and improved it until the accuracy of ASSET became comparable to that of human beings. Furthermore, we developed a spatial and temporal standardization method that allows the segmentation to be compared between different specimens. We illustrate also how the range of application of ASSET can be expanded to automatically

analyze fluorescent time-lapse recordings, for instance to track the centrosomes or the plasma membrane. More generally, ASSET has been designed as a user-friendly Matlab® software that can be readily extended to monitor a variety of cellular processes and thus serves as a standard quantification platform for early *C. elegans* embryos.

## Results

### *Segmentation of the eggshell and the cortex*

The eggshell surrounds and protects the embryo and sets the limits within which early development takes place in *C. elegans*. We decided to take advantage of this natural boundary to standardize both the shape and the size of the embryo. We thus set out to segment the eggshell, which was achieved in three steps.

The first step performs an Initial Elliptical Registration (Fig. S1). Using the best-fitting ellipse of the eggshell obtained using morphological operations (thereafter termed "initial segmentation"; Fig. 2A, magenta line), the original Cartesian image is projected onto an elliptical coordinate system, more amenable to detecting elliptical objects (Fig. 2B).

The second step entails the refined segmentation of the eggshell in the elliptical coordinate system (Fig. 2D). Tracking the eggshell in each row in the elliptical projection delineates an essentially vertical path following the straightest and longest edges in the image (Fig. 2C, 2F). Finding such a sequence of edges can be related to the shortest path problem (Dijkstra, 1959) that can be successfully implemented for image segmentation using dynamic programming (DP; Bellman, 1952; Baggett et al., 2005). The advantage of DP with respect to optimization methods such as "active contour" (Kass et al., 1988) or other iterative methods is that DP ensures global optimality. Moreover, DP finds the global optimum in a time that scales linearly with the number of discrete positions in the contour. In order for the shortest-path to automatically detect the eggshell, we developed a scoring function that embodies the features of the eggshell that are apparent in DIC images (Eq. 1). This allows ASSET to define a single contour that corresponds to the most probable location of the eggshell (Fig. 2D, 2E, 2G). In addition, ASSET computes the probability of the path using the forward-backward algorithm (Rabiner, 1989; Durbin et al., 1998) applied to the same scoring function, thus providing extra

information on the accuracy of the segmentation (Fig. S2A, S2B; Eq. 4). This can serve to reveal alternative segmentation paths (Fig. S2C, arrowheads) or to identify regions of the eggshell with low contrast where the most probable location has been allocated with less certainty (Fig. S2D, arrowhead).

The third step in the segmentation of the eggshell entails the refinement of the single contour by direct least-square fitting (Fig. 2E, 2G; Fitzgibbon et al., 1999). The Refined Elliptical Registration and the segmentation of the eggshell are utilized to facilitate the segmentation of the underlying plasma membrane and cell cortex (for simplicity hereafter referred to collectively as the cortex). The cortex is detected using a similar workflow as for the eggshell. First, the Refined Elliptical Registration is utilized to project the embryo (Fig. 3A). Then, the eggshell is erased from the projected image (Fig. 3B) and the shortest-path algorithm is applied to find the precise position of the cortex (Fig. 3C), using a second scoring function that recognizes features characteristic of the cortex (Eq. 2).

Overall, the above steps enable ASSET to automatically and rapidly segment the eggshell and the cortex in each frame of DIC time-lapse recordings of early *C. elegans* embryos (Fig. 3D, 3E, Fig. S3).

## *ASSET accurately segments DIC images*

We sought to evaluate the precision achieved by the above method of automated segmentation. As we aim at replacing manual segmentation, we compared the performance of ASSET to that of humans. Given that manual segmentation may differ slightly between individuals, segmentation was performed by the four authors and the resulting average value termed the “reference segmentation” (Fig. S4). To address whether ASSET could faithfully replace manual segmentation, we estimated the expected precision of a human being, hereafter referred to as “manual precision”, by averaging the error between the manual segmentations and the reference segmentation.

We first assessed the importance of the shortest-path optimization, by comparing the outcome of DP to the initial segmentation, using the elliptical

coordinate system (Fig. 4A). As shown in Fig. 4B, accuracy was drastically improved both for the eggshell and the cortex by DP, not only by reducing the average error more than two fold, but also by reducing the corresponding standard deviation approximately four fold. This point is especially important as it indicates that ASSET performs robustly in every frame. Despite this improvement, the average error was still significantly larger than manual precision, in particular for the cortex. This was anticipated because the shape of the cortex is far more complex than that of the eggshell, especially early during the cell cycle when surface contractions are numerous (see Movie S1, Fig. 1).

Consequently, we further improved the precision of ASSET by optimizing the parameters in the scoring function (Eq. 1, 2) using machine learning (ML) algorithms. The parameter values that gave the best score (Table S1) were incorporated into ASSET and used as default values hereafter. This brought the precision of ASSET to the level of manual segmentation for the eggshell, and slightly less than that of manual segmentation for the cortex (Fig. 4B). We analyzed whether the remaining inaccuracies in the cortex were specific of some time-points or positions in the embryo. When examining these errors over time (Fig. 4C), we found only a slight increase around the time of pseudo-cleavage and cytokinesis (~-10 min and 0 min, respectively), suggesting that ASSET performs quite uniformly throughout the first cell cycle but has difficulties segmenting deep invaginations. When examining different angular positions around the embryo during the first cell cycle (Fig. 4D), we found that the error exhibits a major peak corresponding to the position of the pseudo-cleavage furrow (~$\pi/2$ rad), as well as other more minor errors. This confirms that deep and narrow invaginations are more difficult to segment for ASSET, as they represent a significant detour for the shortest-path algorithm, which ignores the deepest part of the furrow if the image is not sufficiently contrasted. Despite this inaccuracy, however, ASSET still faithfully detects the cleavage furrow as a somewhat smaller invagination at the correct position and time.

### *Mapping embryos onto an absolute coordinate system*

In addition to automating segmentation, we sought to standardize the quantified data to account for the slight variability that exists between different specimens. Therefore, we developed a new absolute coordinate system that combines spatial and temporal standardization. To this end, we took advantage of the segmentation of the eggshell (Fig. 5A, 5B), projected it onto the elliptical coordinate system (Fig. 5C, 5F), normalized it (Fig. 5D, 5G) and then projected it back onto a Cartesian coordinate system using a common virtual elliptical embryo of dimensions 50µm by 30µm centered at the origin (Fig. 5E). This transformation can then be applied onto the cortex or any cellular structure from this embryo to similarly standardize their position and shape (Fig. 5A, 5E). To synchronize different time-lapse recordings, we used as time zero the onset of cytokinesis, which is faithfully detected by ASSET, since such detection does not rely on the depth of the invagination**.** Overall, with the incorporation of such spatial and temporal standardization, ASSET allows one to compare the data from different time-lapse recordings in an automatic and coherent manner.

### *Using ASSET to track centrosomes in an absolute coordinate system*

As an illustration of how ASSET can be utilized to automate the detection of given proteins or cellular compartments, we used the algorithm to monitor the behavior of centrosomes during the first cell cycle. To this end, we imaged transgenic embryos expressing GFP-TAC-1, which accumulates at centrosomes of one-cell stage embryos (Movie S2; Bellanger and Gönczy, 2003; Le Bot et al., 2003; Srayko et al., 2003). Combining ASSET with the “à trou” spot detection algorithm (OlivoMarin, 2002), we tracked the position of the two centrosomes in time-lapse recordings with subpixel precision, as achieved recently in embryos up to the sixteen-cell stage using 3D stacks (Jaensch et al., 2010). Importantly, the utilization of an absolute reference coordinate system enabled us to precisely compare the different recordings (Fig. 6A). This allowed us to quantify known aspects of

centrosome behavior, including their position along the AP axis (Fig. 6B; Kimura and Onami, 2007), their orientation with respect to the AP axis (Fig. 6C), the inter-centrosome distance (Fig. 6D; Srayko et al., 2003) as well as the micro-movements of centrosomes that had been measured using the movements of pronuclei as a proxy (Fig. S5; Kimura and Onami, 2007). Our analysis provided novel quantitative information about the average behavior of centrosomes. For instance, we found that the centrosome that ends up being positioned most anteriorly migrates around the male pronucleus during the stage of pronuclear meeting at a speed comparable to its velocity during subsequent centration/rotation (3.6 ± 2.1 μm/min versus 3.2 ± 2.9 μm/min). This result raises the possibility that the two processes are driven by the same mechanism. Importantly in addition, this analysis provided information about the variability of centrosome behavior at different times of the cell cycle. For instance, we found that the position of centrosomes along the AP axis is less variable during mitosis than at earlier stages of the cell cycle (Fig. 6B).

In summary, ASSET can monitor the position of centrosomes in live one-cell stage embryos owing to the coherent combination of various recordings. Moreover, we can extrapolate that ASSET could do so also for other organelles of interest.

### *Using ASSET for the automated quantification of surface contractions using a fluorescent membrane marker*

ASSET can also be used to compare complex behavior in the wild-type and given mutant or RNAi conditions in a quantitative manner. As a test case, we set out to compare cortical invaginations in the wild-type (Movie S3) to those in embryos depleted by RNAi of SAPS-1 (Movie S4), a protein associated with the protein phosphatase 6 PPH-6 (Afshar et al., 2010). In *saps-1*(*RNAi*) embryos, surface cortical contractions are observed initially but cease thereafter, and the pseudo-cleavage furrow is absent (Afshar et al., 2010; Movie S4).

Because invaginations are somewhat difficult to segment for ASSET in DIC images, we monitored transgenic embryos expressing a fusion protein between mCherry and a pleckstrin homology (PH) domain targeted to the plasma membrane (Fig. S6A; Kachur et al., 2008). Segmentation using the fluorescent signal was

performed in a similar manner as that of the cortex for the DIC image (Fig. S6B-E), adapting the scoring function to Eq. 3. This enabled us to go beyond the information contained in the DIC images and track more precisely the shape and depth of cortical invaginations over time (Fig. 7A, 7B).

Like for DIC segmentation, we manually tracked the cortex in the fluorescence channel to perform ML on the parameters of Eq. 3, reaching a precision equivalent to that of manual segmentation (Fig. S7A). Furthermore, having manually segmented the DIC and fluorescence channels from the same recordings, we could clearly observe the distortion introduced by DIC optics (Fig. S7B). We sought to correct this distortion and found that there is an intrinsic relationship between the distortion and the illumination of DIC images (Fig. S7B). As a result, we used a simple linear function to reduce the discrepancy between the DIC and fluorescence segmentations (Fig. S7C, S7D). Using this novel empirical correction, we could significantly improve the congruence of DIC segmentation with the fluorescence channel (Fig. S7A, S7E-G).

We then addressed whether using ASSET on embryos expressing mCherry:PH could reveal differences in cortical invaginations between wild-type and *saps-1(RNAi)* embryos. As anticipated, ASSET recapitulated the observations that were reached by manual analysis (Afshar et al., 2010). Thus, ASSET found that *saps-1*(*RNAi*) embryos undergo ruffling during pronuclear formation (Fig. 7C), and that the pseudo-cleavage furrow is absent thereafter (Fig. 7D), whereas subsequent centration/rotation and cytokinesis proceed as in the wild-type (Fig. 7E, 7F). Interestingly, ASSET revealed that not only is the pseudo-cleavage furrow affected, but also that the smaller invaginations during that stage are significantly less deep and cover a smaller area than in the wild-type (ratio of *saps-1*(*RNAi*)/wild-type values: 0.68±0.62 for depth and 0.80±0.59 for area; Student's two-tailed T-test p-value < 0.0001 for both).

Overall, we conclude that ASSET not only detects differences in cortical behavior that can be spotted by manual analysis, but also reveals more subtle quantitative changes that could otherwise go unnoticed.

## Discussion

We developed ASSET as a versatile algorithm for the automated segmentation and standardization of cellular processes in live one-cell stage *C. elegans* embryos. Previous efforts have automated recognition of features in developing *C. elegans* embryos. For instance, automatically monitoring the birth and position of each cell using 4D DIC microscopy or fluorescence microscopy of a histone-GPF fusion protein allows one to follow cell lineages until gastrulation (Schnabel and Priess, 1997; Bao et al., 2006). Such robust automated analysis lead to unexpected conclusions regarding the variability of nematode development, including the importance of cell migration during early embryogenesis (Bischoff and Schnabel, 2006). Software has also been developed to track the position of pronuclei and nuclei using 4D DIC microscopy of early *C. elegans* embryos, which led to novel insights regarding the mechanisms governing centration/rotation (Kimura and Onami, 2007). Similarly the movements of NMY-2 (Non Muscle Myosin -2)-GFP aggregates were automatically tracked from time-lapse recordings during the early stages of centration/rotation, thus clarifying the implication of this motor protein in this process (Goulding et al., 2007). Whilst powerful, the above algorithms do not include a standardization procedure, which might hide aspects of the data. We developed ASSET with the aim of automatically and precisely tracking the cortex and any subcellular structure in a generic manner in early *C. elegans* embryos, reporting the data onto a standardized embryo amenable to further analytical scrutiny.

Classification algorithms have been utilized to undertake an automated phenotypic analysis of *C. elegans* development. Thus, for the early embryo, raw DIC images from time-lapse recordings have been utilized with an energy-based learning network to determine the stage of an embryo up to the 4-cell stage (Ning et al., 2005). Segmentation is performed implicitly in this classification method, which can differentiate in a coarse manner between the nucleus, cytoplasm and external medium. Another approach enabled differentiating between 15 mutant strains that

affect adult morphology and behavior based on 5 min low-magnification time-lapse recordings of adult worms via features based on the segmentation of the body (Geng et al., 2004). In more general terms, since classification usually requires a segmentation step prior to getting quantifiable features, phenotypic classification of *C. elegans* embryos will likely benefit from ASSET, particularly when used in combination with our standardization procedure that leads to highly consistent data sets.

Standardization procedures have indeed proven crucial to thoroughly compare different specimens. For instance, this was essential to create digital atlases of gene expression derived from sampling expression patterns of a large number of genes onto a virtual consensus representation. Thus, standardization of 3D fluorescent images of fixed specimens stained with a DNA marker allowed to analyze *Drosophila* blastoderm embryos (Fowlkes et al., 2008), *C. elegans* larvae (Castro et al., 2009) and early *zebrafish* embryos (Long et al., 2009). However, the complexity of the reconstructed organism and the amount of data imposes that such methods be restricted to fixed specimens. One way to overcome this limitation is to use digital scanned laser light-sheet fluorescence microscopy, which produces highly contrasted images using a short exposure time by imaging whole sections of live specimens at once (Keller et al., 2010). When implementable, this produces data particularly amenable to automated segmentation and standardization, thus leading to detailed quantitative representation of embryogenesis (Keller et al., 2010).

ASSET quantification and standardization capabilities are applicable to a broad range of properties in the early *C. elegans* development. For instance, it allows the straightforward quantification of variability in wild-type *C. elegans* embryos (Fig. S8). Certainly, the scope of applications that can be tackled using ASSET clearly extends beyond the scope of the current manuscript. Moreover, ASSET offers flexibility, because the only assumption for the Initial Elliptical Registration is that the specimen is elliptical. Therefore, ASSET is expected to similarly faithfully segment embryos of other species of nematodes or phyla. In addition, as ASSET segments images indifferently from their origin, it should be able to properly

segment 4D time-lapse recordings, as well as images of embryos fixed and stained with antibodies.

In conclusion, we developed a versatile and robust algorithm for the automated quantification and standardization of *C. elegans* embryos that should contribute to accelerate the pace of discovery by offering a novel tool for systematic and unbiased quantitative studies.

## Experimental Procedures

### *Imaging worm strains*

All *C. elegans* strains were maintained according to standard procedures (Brenner, 1974) and embryos prepared and imaged using DIC time-lapse microscopy at 23 °C, taking 1 image every 10 s (Gönczy et al., 1999). *saps-1*(*RNAi*) was performed as described (Afshar et al., 2010). Transgenic animals expressing mCherry:PH (Kachur et al., 2008) and GFP-TAC-1 (Bellanger and Gönczy, 2003) were grown at 24°C and dissected in osmotically balanced blastomere culture medium (Shelton and Bowerman, 1996). The resulting embryos were imaged using dual time-lapse DIC and fluorescent microscopy on a Zeiss Axioplan 2 with a 6% neutral density filter to attenuate the 103W Arc Mercury epifluorescent source. The motorized filter wheel, two external shutters, and the 1392 x 1040 pixels 12-bit Photometrics CoolSNAP ES[2] were controlled by µManager[1]. Images were taken every 10 s, with an exposure time of 100ms for the DIC and 250ms for the fluorescence channel using the Zeiss Filter Set 10 (GFP) and 43HE (mCherry).

### *Initial elliptical registration and projection*

The approximate position, size and orientation of the embryo in the image were estimated by taking advantage of the rough texture generated by the cytoplasmic yolk granules (Fig. S1A), which results in a much higher density of edges inside the embryo (Fig. S1B). Initial edge detection was performed using the absolute difference mask (ADM) algorithm (Chen and Alzahrani, 1997), and then a threshold applied to maximize the variance between pixels above and below this value, thus retaining only the strongest edges (Fig. S1C; Otsu, 1975). Morphological opening of the image was performed thereafter, thus merging edges that are close to one another and forming a binary region (Fig. S1D), the contour of which was the initial segmentation. Finally, the ellipse that best fits the position, size and orientation of this binary shape was computed using direct least-square fitting (Fig.

S1D, magenta ellipse; Halır and Flusser, 1998), thus providing the Initial Elliptical Registration, which is particularly accurate in the image shown in Fig. 2.

The original Cartesian image (Fig. 2A) was then projected onto an elliptical coordinate system centered and aligned with respect to the embryo, so that each row of the projected image represents a radius of the best fitting ellipse and each column a different radial position (Fig. 2B). The transformations between these two coordinate systems were derived from standard trigonometry (Equation not shown). Because the lack of precision inherent to the approximation might lead to the truncation of the eggshell, which would compromise the subsequent refined segmentation, such potential mistakes were avoided by projecting an area 20% bigger than the estimated ellipse (Fig. 2B).

### *Shortest path algorithm*

The scoring function designed to segment the eggshell (Eq. 1) embodies three principal characteristics: the eggshell is smooth, it is the right-most edge in the image and it is composed of fairly strong edges (Fig. 2C, 2F).

To segment the cortex, the eggshell was erased from the projected image by shifting inwards the area just outside of the embryo (Fig 3B), thus allowing the shortest-path algorithm to detect the cortex without perturbation from the eggshell (Fig. 3C). The scoring function to segment the cortex (Eq. 2) relied on three characteristics: the cortex is locally smooth, it is the outer-most edge of the embryo proper and it is composed of strong and continuous edges; moreover, three terms were added that utilize the anisotropy of DIC images to allow good discrimination between the cell cortex and yolk granules (Eq. 2).

Segmentation was completed using dynamic programming (Dijkstra, 1959), which solves the maximum likelihood path problem recursively, combining at every step the local best solution, provided by the scoring function, to obtain the global optimum. This reduced the segmentation task to finding the best transitions between two consecutive rows in the elliptical projection, thus linking each pixel of a given row with the “best” pixel from the previous row.

### *Manual segmentation*

A dataset was built that accounts for the variability found amongst dual DIC and fluorescence time-lapse recordings of one-cell stage *C. elegans* embryos. 15 frames were randomly chosen from each of 5 time-lapse recordings of one-cell stage embryos, thus obtaining frames from every stage and with different polarizations. The four authors manually segmented these 75 frames, with 2 sets of 15 frames being segmented 3 times, thus measuring also the consistency of an individual in such a task, which did not vary significantly (data not shown).

### *Parameter optimization and machine learning*

The particle swarm optimization (PSO; Kennedy and Eberhart, 1995), an evolution strategy with covariance matrix adaptation (CMA-ES; Hansen and Ostermeier, 2001) and a combination of various population-based optimization schemes (GODLIKE by Rody Oldenhuis[2]) were utilized in parallel. Each iteration of the algorithm segmented all 75 frames to recompute the error (i.e. the average distance to the reference segmentation) using the current set of parameter values. The fitness of the current parameters was computed as the mean error plus one standard deviation (Eq. 5). Including the standard deviation was important to avoid solutions that would be accurate only for some frames. Convergence was reached when the error gradient (i.e. the difference between two consecutive iterations) was less than 1e-25. Overfitting was ruled out by visually inspection of independent frames.

### *Tracking of centrosomes*

The position of the centrosomes was determined by first using the "à trou" spot detection algorithm (OlivoMarin, 2002) and then looking for local maxima greater than three times the mean absolute deviation of the correlation image. Thereafter a Gaussian was fitted around this position, thus reaching subpixel precision, and overlapping detections merged (Jaqaman et al., 2008). Ultimately, the two brightest detected Gaussians were kept. Centrosomes were tracked between frames by globally minimizing the distance that the two entities will travel (Jaqaman

et al., 2008), linearly interpolating the missing positions using the previous and following frames (see Movie S2).

### *Segmentation using fluorescence*

mCherry:PH images (Fig. S6A) were preprocessed using a combination of Gaussian and median filtering to increase signal-to-noise ratio (Fig. S6C), projected (Fig. S6B) as for DIC images and segmented using the shortest-path algorithm (Fig. S6C) using a simpler scoring function (Eq. 3) because the cortex is characterized by a fairly smooth succession of the outer-most bright pixels. More naïve segmentation methods were tested but were less precise (data not shown).

### *Quantification of the invaginations*

Since cortical invaginations are located in concave areas, the convex hull of the segmentation of the cortex was computed and its distance to the actual segmentation measured. The peaks in elliptical coordinates correspond to invaginations. A heuristic threshold of 2% was applied to filter the detection of local maxima. For each invagination, the depth and the area were quantified and their duration determined over consecutive frames as described (Jaqaman et al., 2008).

To orient the embryo along the AP axis (i.e. to ensure that the posterior pole is located at 0 rad; Fig. 4A), the position of the cleavage furrow was utilized, because cytokinesis invariably occurs in a slightly posterior location in the wild-type. The cytokinesis furrow was backtracked from the last frame of the recording until the first frame in which it was detected, which was set as time 0.

### Distortion between DIC and fluorescence segmentation

The correction to apply between DIC and fluorescence segmentation was determined by measuring the distance between the corresponding manual segmentations. As the correction has a sinusoidal shape close to that of the pixel intensity around the eggshell ($I$; Fig. S7B), a linear regression was used to correlate them. An additional angular shift was introduced to account for the one that is observed heuristically (Fig. S7B, arrow). To avoid depending on the contrast of the image, the pixel intensities of each image was scaled independently and a term

accounting for the width of this range added to the fit ($R$). To identify the best angular shift, all possible values were exhaustively tested with a step of π/100. A linear regression between the measured eggshell pixel intensity and each shifted correction was performed and the performance of the fit measured using the root mean squared error (Fig. S7C). The best parameters were as follows: shift 5.1836 rad, linear fit 0.0424 - 0.0440*$I$ - 0.0176*$R$ (Fig. S7D). This reduced the segmentation error from 0.042 ± 0.031 % of the embryo's radius (ER) to 0.019 ± 0.017 % of ER (Fig. S7E-G).

**Posterior Decoding**

To apply the forward-backward algorithm (Durbin et al., 1998) to our data, the scoring functions were converted into the corresponding emission (i.e. to be part of the segmentation) and transition (i.e. to move to a given pixel) probabilities using Eq. 4. The conversion has two unspecified parameters that represent the sharpness of the probability distribution (i.e. the exponent of the distribution).

The exponent for the transition probabilities (λ) was first fixed independently from that for the emission probabilities (μ), using the fact that with uniform emission probabilities (i.e. no information from the image), the paths would behave like a random walk. λ was fixed such that the expected translation distance at the end of the random walk (i.e. in the last row of the image) equals W/6 (W: width of the image). λ was fixed by dichotomy using $P_{trans}$ and an initial value of 1. μ was fixed by setting the mean standard deviation of the posterior probability to W/50, ensuring a fairly peaked posterior probability from which information can still be extracted. The corresponding value of μ was set by dichotomy using an initial value 1 and the value previously found for λ.

## *Software implementation*

The algorithm has been tested on Matlab® R2009b, is executed in ~5 sec per frame and is freely available upon request. Manual segmentation was performed using the ImageJ plug-in A_3D_editing from Albert Cardona[3]. The ML algorithms were done using the Matlab® implementation of the publicly available algorithms:

CMA-ES[4] by Nikolaus Hansen, PSO[5] by Brian Birge, GODLIKE[6] by Rody Oldenhuis. The user-friendly graphical interface was based on TSPACKGUI[7] by Robert Renka.

### *Abbreviations*

AP: Anteroposterior; ASSET: Algorithm for the segmentation and the standardization of *C. elegans* time-lapse recordings; DIC: Differential interference contrast**;** ER: Embryo radius**;** DP: Dynamic programming; ML: Machine learning; PH: Pleckstrin homology; SAPS-1: Sit4p-associated proteins-1; TAC-1: Transforming acidic coiled-coil-1.

## Acknowledgments

We are grateful to the association "Josy Marty – Echec au Cancer de la Broye", the Swiss National Science Foundation and SystemsX.ch for financial support, to Anjon Andhya for transgenic animals expressing mCherry:PH, as well as to Michael Unser, Jonathan Bieler, Teresa Ferraro, Aude Meier and Virginie Hamel Hachet for useful comments on the manuscript. We also would like to thank Albert Cardona, Nikolaus Hansen, Brian Birge, Rody Oldenhuis, Robert Renka, Nikolai Chernov, Bolu Ajiboye, Jonas Dorn, Pekka Kumpulainen, Yi Cao, Juerg Schwizer, Jochen Lenz and the Bio-Formats[8] for useful pieces of code. We are thankful to the administration team of the Greedy[9] cluster for enabling ML to be computed on this grid.

## References


Afshar K, Werner ME, Tse YC, Glotzer M, Gönczy P. 2010. Regulation of cortical contractility and spindle positioning by the protein phosphatase 6 PPH-6 in one-cell stage *C. elegans* embryos. Development (Cambridge, England) 137:237-247.

Baggett D, Nakaya M-a, McAuliffe M, Yamaguchi TP, Lockett S. 2005. Whole cell segmentation in solid tissue sections. Cytometry. Part A : the journal of the International Society for Analytical Cytology 67:137-143.

Bao Z, Murray JI, Boyle T, Ooi SL, Sandel MJ, Waterston RH. 2006. Automated cell lineage tracing in *Caenorhabditis elegans*. Proceedings of the National Academy of Sciences of the United States of America 103:2707-2712.

Bellanger JM, Gönczy P. 2003. TAC-1 and ZYG-9 form a complex that promotes microtubule assembly in *C. elegans* embryos. Current biology : CB 13:1488-1498.

Bellman R. 1952. On the Theory of Dynamic Programming. Proceedings of the National Academy of Sciences of the United States of America 38:716-719.

Bischoff M, Schnabel R. 2006. Global cell sorting is mediated by local cell-cell interactions in the *C. elegans* embryo. Developmental biology 294:432-444.

Brenner S. 1974. The genetics of Caenorhabditis elegans. Genetics 77:71.

Castro C, Luengo-Oroz Ma, Desnoulez S, Duloquin L, Fernandez-de-Manuel L, Montagna S, Ledesma-Carbayo MJ, Bourgine P, Peyrieras N, Santos a. 2009. An automatic quantification and registration strategy to create a gene expression atlas of *zebrafish* embryogenesis. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference 2009:1469-1472.

Chen T, Alzahrani FM. 1997. A Real-Time Edge Detector: Algorithm and VLSI Architecture. Real-Time Imaging 3:363–378.

Dijkstra EW. 1959. A note on two problems in connexion with graphs. Numerische Mathematik 1:269-271.

Durbin R, Eddy S, Krogh A, Mitchison G. 1998. Hidden Markov models. In: Biological sequence analysis. pp 52 - 62.

Fitzgibbon a, Pilu M, Fisher RB. 1999. Direct least square fitting of ellipses. IEEE Transactions on Pattern Analysis and Machine Intelligence 21:476-480.

Fowlkes CC, Hendriks CLL, Keränen SVE, Weber GH, Rübel O, Huang M-Y, Chatoor S, DePace AH, Simirenko L, Henriquez C, Beaton A, Weiszmann R, Celniker S, Hamann B, Knowles DW, Biggin MD, Eisen MB, Malik J. 2008. A quantitative spatiotemporal atlas of gene expression in the *Drosophila* blastoderm. Cell 133:364-374.

Geng W, Cosman P, Berry CC, Feng Z, Schafer WR. 2004. Automatic tracking, feature extraction and classification of *C. elegans* phenotypes. IEEE transactions on bio-medical engineering 51:1811-1820.

Gönczy P. 2008. Mechanisms of asymmetric cell division: flies and worms pave the way. Nature reviews. Molecular cell biology 9:355-366.

Gönczy P, Schnabel H, Kaletta T, Kaletta T, Amores AD, Hyman T, Amores AD, Hyman AA, Schnabel R. 1999. Dissection of cell division processes in the one cell stage *Caenorhabditis elegans* embryo by mutational analysis. The Journal of cell biology 144:927-946.

Goulding MB, Canman JC, Senning EN, Marcus AH, Bowerman B. 2007. Control of nuclear centration in the *C. elegans* zygote by receptor-independent Galpha signaling and myosin II. The Journal of cell biology 178:1177-1191.

Halır R, Flusser J. 1998. Numerically stable direct least squares fitting of ellipses. The Sixth International Conference in Central Europe.

Hamahashi S, Onami S, Kitano H. 2005. Detection of nuclei in 4D Nomarski DIC microscope images of early *Caenorhabditis elegans* embryos using local image entropy and object tracking. BMC bioinformatics 6:125.

Hansen N, Ostermeier A. 2001. Completely derandomized self-adaptation in evolution strategies. Evolutionary computation 9:159-195.

Jaensch S, Decker M, Hyman Aa, Myers EW. 2010. Automated tracking and analysis of centrosomes in early *Caenorhabditis elegans* embryos. Bioinformatics (Oxford, England) 26:i13-20.

Jaqaman K, Loerke D, Mettlen M, Kuwata H, Grinstein S, Danuser G. 2008. Robust single-particle tracking in live-cell time-lapse sequences. Nature Methods 5.

Kachur TM, Audhya A, Pilgrim DB. 2008. UNC-45 is required for NMY-2 contractile function in early embryonic polarity establishment and germline cellularization in *C. elegans*. Developmental biology 314:287-299.

Kass M, Witkin A, Terzopoulos D. 1988. Snakes: Active contour models. International Journal of Computer Vision:321-331.

Keller PJ, Schmidt AD, Santella A, Khairy K, Bao Z, Wittbrodt J, Stelzer EHK. 2010. Fast, high-contrast imaging of animal development with scanned light sheet-based structured-illumination microscopy. Nature methods 7:637-642.

Kennedy J, Eberhart R. 1995. Particle Swarm Optimization. Proceedings of IEEE International Conference on Neural Networks 4:1942-1948.

Kimura A, Onami S. 2007. Local cortical pulling-force repression switches centrosomal centration and posterior displacement in *C. elegans*. The Journal of cell biology 179:1347-1354.

Le Bot N, Tsai MC, Andrews RK, Ahringer J. 2003. TAC-1, a regulator of microtubule length in the *C. elegans* embryo. Current biology : CB 13:1499-1505.

Long F, Peng H, Liu X, Kim SK, Myers E. 2009. A 3D digital atlas of *C. elegans* and its application to single-cell analyses. Nature methods 6:667-672.

Ning F, Delhomme D, LeCun Y, Piano F, Bottou L, Barbano PE. 2005. Toward automatic phenotyping of developing embryos from videos. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society 14:1360-1371.

Oegema K, Hyman Aa. 2006. Cell division. WormBook : the online review of *C. elegans* biology:1-40.

OlivoMarin J. 2002. Extraction of spots in biological images using multiscale products. Pattern Recognition 35:1989-1996.

Otsu N. 1975. A threshold selection method from gray-level histograms. Automatica 11:285–296.

Rabiner LR. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77:257-286.

Schnabel R, Priess JR. 1997. Specification of cell fates in the early embryo. In: *C. elegans* II. Cold Spring Harbor Laboratory Press. pp 361-382.

Shelton Ca, Bowerman B. 1996. Time-dependent responses to glp-1-mediated inductions in early *C. elegans* embryos. Development (Cambridge, England) 122:2043-2050.

Srayko M, Quintin S, Schwager A, Hyman AA. 2003. *Caenorhabditis elegans* TAC-1 and ZYG-9 Form a Complex that Is Essential for Long Astral and Spindle Microtubules. Current Biology 13:1506-1511.

# Figures legends

## Figure 1

Early events in one cell-stage *C. elegans* embryo

**(A-D)** Images from a time-lapse DIC microscopy recording of a one-cell stage *C. elegans* embryo from the exit from meiosis II until the end of the first mitosis (see Movie S1). In this and other figures, embryos are represented with anterior on the left and posterior on the right, while scale bars correspond to 10µm. Time to cytokinesis onset is depicted in minutes in the top right corner of each frame.

**A** Pronuclear formation and ruffling of the cell cortex.

**B** Pronuclear migration during the pseudo-cleavage stage.

**C** Centration/rotation of the two pronuclei and associated centrosomes, as well as mitosis. Note that one pronucleus is out of focus in this particular frame.

**D** Cytokinesis; arroweads point to ingressing cleavage furrow.

**(E-H)** Schematic representations corresponding to (A-D). Eggshell in blue, plasma membrane/cortex in red, DNA in purple, centrosomes in orange and microtubules in green. Note that neither microtubules nor centrosomes can be clearly observed by DIC before mitosis.

**Figure 2**

Elliptical projection and automated segmentation of the eggshell

**A** Raw DIC image overlaid with the Initial Elliptical Registration (magenta) and its center (red dot); numbers in panels A and B denote corresponding locations.

**B** Elliptical projection of A; the red line (left) and the magenta line (right) denote the center and the perimeter of the Initial Elliptical Registration, respectively. Note that the area on the right of the magenta line corresponds to the additional zone that is projected to prevent truncation of the eggshell.

**C** Edges detected in B by the ADM filter; note that the brightness and contrast were adjusted for better visibility. The white box delimits the area magnified 6 folds in F; note that all magnified views use this factor.

**D** Automated segmentation of the eggshell (green) on the projection.

**E** Outcome of the segmentation (green) converted back onto the original image with the re-estimated best-fitting ellipse (cyan). Note that the Initial Elliptical Registration (magenta) is also drawn but covered by the new best-fitting ellipse in this particular frame. The white box delimits the area magnified in G.

**F** Magnification of the area delimited by a white box in C; the segmented eggshell (delimited by the green brackets) is fairly straight, vertical and displays a continuous edge. Note that the reference eggshell derived from manual segmentation was usually located halfway between the outer and the inner edges of the eggshell.

**G** Magnification of the area delimited by a white box in E.

## Figure 3

Elliptical projection and automated segmentation of the cortex

**A** Elliptical projection of the embryo (similar to Fig. 2B) overlaid with the detected inner and outer part of the eggshell (which appear to be overlapping at this low magnification).

**B** Elliptical projection after removal of the eggshell.

**C** Result of the automated segmentation of the cortex (orange) with the previously segmented eggshell (green).

**D** Segmented cortex (orange) and eggshell (green) converted back onto the original DIC image overlaid with the initial segmentation of the cortex (brown, displayed thicker for visual purposes). The white boxes delimit the area magnified in E (upper one), Fig. S3A (lower one) and S3B (middle one).

**E** Magnification of the area delimited by the lower white box in D, showing all three segmentations (green: eggshell; orange: cortex; brown: initial segmentation of the cortex). Note that the reference cortex derived from manual segmentation was typically located on the outermost edge of the cortex.

## Figure 4

Assessing the performance of the automated segmentation

**A** Schematic of the projection between the Cartesian coordinate system of the image and the elliptical coordinate system (ER: embryo radius). Depicted are the eggshell (green) and the cortex (orange) overlaid with points of reference at particular angular positions (purple dots). The elliptical coordinate system is used to measure the performances of ASSET in B-D.

**B** Error of the different segmentation strategies. Each result is represented as the average distance to the reference segmentation, in % of the ER, with its standard deviation. The results are color-coded as indicated and grouped by segmentation type (eggshell: left, cortex: right). The black circle shows the only pair of values not significantly different (Student's one-tailed T-test, p-value = 0.988).

**C** Error of the automated segmentation of the cortex over time. The x-axis shows the time in minutes with t=0 being the onset of cytokinesis. The y-axis is as in B. The asterisks show errors that are significantly larger than the error in B (Student's one-tailed T-test, * p-value < 0.05, ** p-value < 0.0001).

**D** Error of the automated segmentation of the cortex around the embryo. The x-axis shows the error at 36 different angular positions, starting and ending at the anterior pole, rotating counter-clockwise around the embryo. The y-axis is as in B. Asterisks are as in C.

**Figure 5**

Standardization of an embryo

**A** Automated segmentation of the eggshell (green) and the cortex (orange).

**B** Corresponding elliptical projection.

**C** Segmentation in elliptical coordinates. The x-axis is the radial position, the y-axis the angular position. The box delimits the area magnified in F.

**D** Result of the standardization in elliptical coordinates. The box delimits the area magnified in G.

**E** Normalized segmentation of the cortex onto the 50µm by 30µm reference embryo.

**F-G** Magnification of the areas delimited by the boxes in C and D, respectively. Note how the eggshell was straightened after the standardization procedure (G).

## Figure 6

Automated tracking of centrosomes

**A** Overlay of tracking, in the absolute coordinate system, of the two centrosomes in time-lapse recordings of five embryos expressing GFP-TAC-1 (thin lines), from pronuclear formation until cytokinesis onset. Bold lines: average trajectories of anterior (red) and posterior (blue) centrosomes, with the evolution over time color-coded from dark (early) to light (late).

**B** Kymograph of centrosome positions along the AP axis (thin lines: individual recordings; bold: averages). Averages are color-coded as in A, the light areas represent standard deviations. The two dashed lines and the intervening arrow in panels B-D denote the time within which centration/rotation takes place (-6 min to -2 min 50 sec). The numbers displayed are the average speed of movement (in μm/min ± standard deviation) before, during and after centration/rotation for the anterior (red) and posterior (blue) centrosomes.

**C** Kymograph of the angular orientation between the two centrosomes with π representing the correct final alignment along the AP axis (thin lines: individual recordings; bold: average). Time evolution of the average ranging in color from dark (early) to light (late); light areas represent standard deviations from the average. Numbers are average speeds in rad/min ± standard deviation.

**D** Kymograph of the inter-centrosome distance. The representation is the same a C. Average speeds in μm/min ± standard deviation.

## Figure 7

ASSET-based analysis of cortical behavior in wild-type and *saps-1*(*RNAi*) embryos

**A** Automated detection of invaginations in embryos expressing mCherry:PH. The eggshell (green) was segmented using the corresponding DIC image and corrected (Fig. S7) while the cortex (orange) was segmented with the mCherry signal. The red circles represent the position of the detected invaginations.

**B** Comparison of invaginations detected by DIC (gray) and by mCherry:PH (blue) during the pseudo-cleavage stage (-16 min 10 sec to -6 min 20 sec with respect to cytokinesis onset, Fig. 1B). Each dot represents an invagination in any of the frames during this time-span. 10 time-lapse recordings for each condition were analyzed. The solid lines depict the average area of invaginations along the angular position.

**(C-F)** Distribution of invaginations in angular positions at the indicated stages: C, pronuclear formation, -18 min to -16 min 10 sec; D, pseudo-cleavage, -16 min 10 sec to -6 min 20 sec; E, centration/rotation as well as mitosis, -6 min 20 sec to 0; G, cytokinesis, 0 to +1 minutes. Blue: wild-type embryos, red: *saps-1*(*RNAi*) embryos. The absence of pseudo-cleavage in *saps-1*(*RNAi)* embryos can be clearly observed in D (at angular positions $\pi/2$ and $3\pi/2$). The angular areas depicted in gray were excluded for the numerical analysis. Note also that cytokinesis proceeds normally in *saps-1*(*RNAi)* embryos (F).

# Equations

## Equation 1

Scoring function used to segment the eggshell in DIC images.

$$\sum_{ij} = \alpha S(i,j) + (1-\alpha)D(j)$$
$$S(i,j) = \beta\left(\gamma|i-j| + (1-\gamma)|dir(i) - dir(j)|\right) + (1-\beta)|I(i) - I(j)|$$
$$D(j) = \eta(1-E(j))^{\varepsilon} + (1-\eta)\sum_{k\le j} E(k)^{\varepsilon}$$

*I(i)* represents the intensity of pixel *i* and *E(j)* the strength of the edge detected by the ADM filter on pixel *j*. The score $\sum_{ij}$ is computed for a transition from pixel *i* to pixel *j*. The function is a sum of a term that accounts for the smoothness of the path (*S(i,j)*) and another term that accounts for the information contained in the image (*D(j)*). The smoothness penalizes curved paths with one term for the absolute distance between the two pixels, one for the difference in direction, and a third for the difference in intensity between pixels. The information has one term that favors strong edges and one that accounts for edges on the right of pixel *j*. In total, the scoring function has 5 parameters ($\alpha,\beta,\gamma,\eta,\varepsilon$). Without loss of generality we took all of the parameters in the range [0, 1] and normalized the terms of the equation.

## Equation 2

Scoring function used to segment the cortex in DIC images.

$$D(j) = \eta\left(\nu(1-E(j))^{\varepsilon} + (1-\nu)gap(j)\right) + (1-\eta)\left(\psi\sum_{k>j} E(k)^{\varepsilon} + (1-\psi)\left(\xi(1-|I(j)-I_{out}|) + (1-\xi)|I(j)-I(g)|\right)\right)$$

$$gap(j) = \begin{cases} 1, & E(j) = 0 \\ 0, & E(j) \neq 0 \end{cases}$$

$\sum_{ij}$ and *S(i,j)* are the same as defined in Eq. 1. In addition to the two terms used for the detection of the eggshell, three terms have been added: one to penalize gaps (tuned by $\nu$), one that favors pixels with an intensity far from the average outside intensity ($I_{out}$; $\xi$) and one that favors pixels with an intensity comparable to the one of the eggshell ($g$ is the index of the pixel of the eggshell; tuned by $(1-\xi)$). These terms take advantage of the inherent features of DIC polarization, which changes the intensity of the pixels with their angular position. For instance, in Fig. 3B, the cortex is gray at the anterior pole, white on the top, gray again on the posterior pole and black on the bottom. The intensity of the eggshell *I(g)* varies in a similar manner. In contrast, yolk granules are small punctae with a mixture of dark and light pixels throughout the cytoplasm (Fig. 3B). Moreover, the cortex is a continuous edge, whereas yolk granules are separated and thus sensitive to the gap penalty. This function has 8 parameters in total ($\alpha,\beta,\gamma,\eta,\varepsilon,\nu,\psi,\xi$) and each term is normalized as in Eq. 1.

**Equation 3**

Scoring function used to segment the cortex in fluorescence images.

$$D(j) = \eta\left(1 - I(j)\right) + (1-\eta)\sum_{k>j} I(k)$$

$\sum_{ij}$, *S(i,j)* and *I(j)* are the same as defined in Eq. 1. This scoring function has one term that favors light pixels and a second one that penalizes having light pixels on the right of pixel *j*. This function has 4 parameters in total ($\alpha,\beta,\gamma,\eta$) and each term is normalized as in Eq. 1.

**Equation 4**

Computing the posterior decoding probabilities

$$P_{trans}(i,j) = \exp\left(-\lambda\alpha\beta\gamma\left|i-j\right|\right)$$
$$P_{emis}(i,j) = \exp\left(-\mu\alpha\left(\beta(1-\gamma)\left|dir(i)-dir(j)\right| + (1-\beta)\left|I(i)-I(j)\right|\right) + (1-\alpha)D(j)\right)$$

The conversions use the parameters from the scoring functions (Eq. 1, 2, 3; i.e. the maximum likelihood) but have two unspecified standardization constants (λ, μ), which need to be set in order to use posterior decoding.

**Equation 5**

The fitness function used for ML.

$$f(p) = \overline{E(p)} + \sqrt{\frac{1}{N}\sum_{n}\left(E(n,p) - \overline{E(p)}\right)^2} + \exp\left(\sum_{i} B(p_i)\right)$$

$$B(p_i) = \begin{cases} L_i - p_i & p_i < L_i \\ p_i - U_i & p_i > U_i \\ 0 & L_i \le p_i \le U_i \end{cases}$$

*p* is a vector of the parameters of the scoring function, *E(n,p)* the error of the segmentation of frame *n* using *p*, $\overline{E(p)}$ the mean over the different frames, $\sqrt{\ldots}$ the standard deviation of the error and $B(p_i)$ is used to keep the parameters between the numerical bounds $[L_i, U_i]$ for parameter *i*.

## Footnotes

[1] http://www.micro-manager.org
[2] http://www.mathworks.com/matlabcentral/fileexchange/24838
[3] http://www.mcdb.ucla.edu/Research/Hartenstein/software/imagej/
[4] http://www.lri.fr/~hansen/cmaesintro.html
[5] http://www.mathworks.com/matlabcentral/fileexchange/7506
[6] http://www.mathworks.com/matlabcentral/fileexchange/24838
[7] http://www.mathworks.com/matlabcentral/fileexchange/18739
[8] http://www.loci.wisc.edu/software/bio-formats
[9] http://greedy.epfl.ch/

# Supplementary data

## *Table S1*

Value of parameters used in the scoring functions (Eq. 1, 2, 3) after ML optimization.

| Parameter | Eggshell (Eq. 1) | Cortex (Eq. 2) | Fluorescence (Eq. 3) |
|---|---|---|---|
| N | 5 | 5 | 9 |
| $\alpha$ | 0.8873 | 0.7773 | 0.3336 |
| $\beta$ | 0.1813 | 0.0769 | 0.2941 |
| $\gamma$ | 0.9975 | 0.3467 | 0.6278 |
| $\eta$ | 0.2490 | 0.2090 | 0.5228 |
| $1/\varepsilon$ | 0.0240 | 0.6979 | Ø |
| $\nu$ | Ø | 0.9854 | Ø |
| $\psi$ | Ø | 0.5326 | Ø |
| $\xi$ | Ø | 0.8751 | Ø |

# Figures legends

## *Figure S1*

Computation of the Initial Elliptical Registration for the DIC segmentation

**A** Raw DIC image.

**B** Edges detected by the ADM filter; note that the brightness and contrast were adjusted for better visibility. .

**C** Thresholded edges.

**D** Fusion of the detected edges through morphological opening, overlaid with the best fitting ellipse (magenta).

## *Figure S2*

Posterior Decoding for probabilistic segmentation

**A** Projection of the original DIC image overlaid with the maximum-likelihood segmentation for the eggshell (green) and the cortex (orange).

**B** Posterior probability distribution underlying the segmentation of the eggshell on the same projection. The values of the probabilities are color-coded as indicated by the color bar. The white boxes delimit the areas magnified in C (upper one) and D (lower one).

**C** Magnification of the area delimited by the top white box in B. The white arrowheads point to an example of multiple locally optimal segmentation paths.

**D** Magnification of the area delimited by the white box in B. The arrowhead points to a portion of lower confidence (lower probability) in the segmentation.

## *Figure S3*

Automated segmentation for different polarization of the light

**A** High magnification of an area with bright illumination delimited by the white box on the top in Fig. 2D.

**B** High magnification of an area with intermediate illumination delimited by the white box on the right in Fig. 2D.

## *Figure S4*

Comparison between manual and automated segmentation

**A** High magnification of the area delimited by the white box in Fig. 2E.

**B** Three different manual tracking by individual A (Tracking #1, #2 and #3; turquoise: eggshell; purple: cortex), with their respective averages (Manual #A, cyan/pink).

**C** The four averages segmentations (Manual #A, #B, #C and #D) for the eggshell/cortex (cyan/pink), overlaid with their respective average (blue/red), i.e. the reference manual segmentation.

**D** Comparison between the reference manual segmentation (blue/red), the results of ASSET (green/orange) and the initial segmentation (dark green/brown).

**E** Magnification of the area delimited by the white box in Fig. 2D and which corresponds approximately to the elliptical projection of A. Yellow: area measured to compute the error between the reference segmentation (red) and the automated one (orange) (i.e. the area between both segmentations divided by the length of the reference); the error is 0.0195 % of the ER (< 0.0049 µm); white: area between the reference and the segmentation without DP (brown), the error in this case is 0.0289 % (< 0.0072 µm).

## *Figure S5*

Additional analysis of centrosomes micro-movements

Each dot represents the movement between two frames (10 sec interval) of the anterior (red dots) and the posterior (blue dots) centrosome. The thin dots are from individual recordings, the bold ones depict movements of the average centrosomes. These movements are recorded from pronuclear formation until pronuclear meeting (i.e. -12 min to -6 min; left panel), during centration/rotation (-6 min to -2 min 50 sec; middle panel), as well as during mitosis (-2 min 50 sec to 50 sec; right panel). Duration of centration/rotation was determined from the movies using the speed of rotation of the centrosomes, as it is bounded by two periods of very low rotation speed (< 0.02 rad/min). As in (Kimura and Onami, 2007), we observe a bias towards the anterior side of the embryo in the first two panels, while the oscillations of the spindle observed during mitosis are very obvious in the bottom panel as displacements along the Y axis.

## *Figure S6*

Segmentation from fluorescence images

**A** Original mCherry::PH image.

**B** Elliptical projection of the processed image (removed background, Gaussian and median filtering) overlaid with the result of the segmentation for the cortex (orange) and the eggshell (green). The white box delimits the area magnified in D.

**C** Result of the segmentation of the cortex (orange) and the eggshell (green) with the re-estimated best fitting ellipse (cyan) on top of the processed image. The white box delimits the area magnified in E.

**D**-**E** Magnification of the area delimited by the box in B and C, respectively.

## *Figure S7*

Correcting the distortion of DIC with respect to fluorescence imaging

**A** Error of the different segmentation strategies for the mCherry:PH channel. Each result is represented as the average distance to the reference segmentation, in % of the embryo's radius (ER), with its standard deviation. The results are color-coded as indicated. The circle shows the only pair of results that are not significantly different (Student's one-tailed T-test, p-value = 0.7337).

**B** Comparison of the distortion introduced by DIC (red) and the pixel intensity on the eggshell (blue). Both signals have a similar sinusoidal shape but they are displaced by an angular shift (black arrow). The thin lines are data from a particular movie while the thick ones are their averages.

**C** Fitness of the regression for different values of the shift between the pixel intensity profile and the correction (x-axis). The y-axis represents the precision of the correction (root mean squared error of the linear fit). The minimum is depicted by the red circle, at position 5.18 rad.

**D** Linear fit performed on the correction. Each black dot represents the pixel intensity (x-axis) with respect to the value of the correction (y-axis). The red line is the result of the linear regression with its width representing its standard deviation. $R^2$ is the coefficient of determination of the regression.

**E** Example of correction for the cortex with the DIC segmentation (light orange), the fluorescence segmentation (dark orange) and the correction (medium orange). The error with the DIC segmentation is 0.0440 % of the ER (< 0.011 µm) while the one after correction is 0.0217 % of the ER (< 0.0054 µm). The white box delimits the area magnified in F and G.

**F-G** Magnification of the areas delimited by the white boxes in E. Note how the DIC segmentation is closer to the fluorescence one after correction (F); note also the large discrepancy due to the polar body which can be seen on the mCherry:PH channel but not on the DIC one (G).

## *Figure S8*

Variability in size and shape of wild-type embryos

**A-E** Quantification of several properties of the ten wild-type embryos studied in Fig. 7. Each panel displays measurements from all the frames of every movie, depicted as boxplots, were the red line is the median, the box ranges from the 25$^{th}$ to the 75$^{th}$ percentile, the bars reach the position of the most extremes data points and the crosses represent outliers. Note that the range of the y-axis is different for each panel. Note also that part of the observed variability might be due to changes of the focal plane.

**A** Length of the major radius of the refined elliptical registration. Note that the major radius of our reference embryo measures 25µm.

**B** Length of the minor radius. Note that the minor radius of our reference embryo measures 15µm.

**C** Ratio of the major versus the minor radii. Given a 50µm by 30µm ellipse, this ratio would be 1.67.

**D** Total length of the cortex. Given an ellipse that has the properties of the median from A and B, the total length of the cortex perimeter should be ~148µm. Note that the outliers correspond to embryos in which the recordings lasted longer, allowing for deeper cleavage furrows to be tracked, thus increasing significantly the total length of their cortex.

**E** Cytoplasmic area enclosed by the cortex. Given a 50µm by 30µm ellipse, the cytoplasmic area would be ~1590µm$^2$.

## Movies Legend

Movies S1: Time-lapse DIC microscopy of wild-type one-cell stage *C. elegans* embryo, along with the resulting segmentation by ASSET (green: eggshell, orange: cortex). All movies are accelerated 60 times compared to the actual pace of development; anterior is on the left, embryos are ~50 μm-long and time is indicated in minutes and seconds with respect to cytokinesis onset.

Movie S2: Time-lapse fluorescence recording of transgenic one-cell stage *C. elegans* embryo expressing GFP-TAC-1 and resulting tracking of centrosomes overlaid with the corrected segmentation performed in the DIC channel (red circle: anterior centrosome, blue circle: posterior centrosome, green: eggshell, orange: cortex).

Movie S3-S4: Time-lapse fluorescence recording of wild-type (Movie S3) or *saps-1*(*RNAi*) (Movie S4) transgenic one-cell stage *C. elegans* embryos expressing mCherry::PH and resulting detection of invaginations (purple dots: invagination, green: eggshell from the DIC channel, orange: cortex).